Computerized Support for Research and Publication in Contemporary History

Zoltán Lux

Abstract: The "1956 Institute" deals with research regarding Hungarian history since the Second World War with an emphasis on the 1956 Hungarian Revolution and its development, subsequent effects and international aspects. Since its foundation in 1990, the institute has been amassing databases containing the documents gathered and used in the researches or descriptions of these documents.

After briefly introducing the "1956 Institute", the paper will focus on the following problems in the process of data compilation and provision: 1) Databases at the "1956 Institute: oral history interviews, trial records and the photographic database; 2) Establishment of a new multimedia data-base; 3) The problem of sharing individually possessed information, and of disposal and control over information, 4) Compiling digital publications.

Table of Contents

1. The "1956 Institute"

The "1956 Institute" was founded in 1990 to conduct research on Hungarian history from the period of Second World War and on, including the international aspects of it. The primary subject of this research was the Hungarian Revolution of 1956. However, this subject has been gradually expanded over the years, covering the time span from the end of Second World War to the collapse of the state-socialist system. [1]

The "1956 Institute" is not an archive of social science data as such. Sometimes its researchers utilize data from other archives and libraries, and sometimes they act as "data providers" themselves. The various types of data and documents gathered during the research work allow the re-use, partly or wholly, in

subsequent researches: Data can be placed in a broader context, refined, augmented or possibly refuted. There is nothing essentially new in this process, except that on the path towards an information society, far more data and documents become accessible quicker than before. Furthermore, the pace of the researches would become faster than previously. [2]

2. Databases at the "1956 Institute"

As the Institute slowly began to build up its information system ten years ago, the focus of attention was on how computer technologies would offer entirely new opportunities in scientific research, especially for historians and sociologists. Word-processing and table-compiling programs were certainly convenient, but they did not constitute the real innovation brought on by the spread of informatics. [3]

The essential change—technically and towards content—came about in 1991, when the Institute began to establish and develop its databases. The intention, within the scope available at that time was to incorporate all the information—data and documents found by the Institute staff during their research—as detailed as possible. That, essentially, would make all the data accessible and available to others, including interim results and details discovered, even if it were not published in full. [4]

The first step was to add to the Institute´s conventional bibliographical database, the historical data that would be extremely useful in more detailed historical research. This meant recording the events and persons mentioned in a book or article and giving a detailed description of each event (date, place, exact location, institutions, side-events, participants), and beginning to gather biographical details on the persons involved and data on other related documents and so on. [5]

It was soon being suggested that it would be worth "abstracting" other types of document used in the researches, in other words filling the database with well-structured, detailed data. This seemed worthwhile with the thousand or more life interviews possessed by the Institute. [6]

2.1 The database for oral history interviews

The researches of the Oral History Archive at the "1956 Institute" cover people who were prominent in the recent history of Hungary. They include *eye-witnesses* to the 1956 Hungarian Revolution—those who influenced and participated in its events as well as its victims—and notable figures in the Hungarian arts and culture, economy, science, academia and public life. The archive holds about a thousand life interviews recorded on tape and transcribed. They follow the methods of Oral History or a sociological life story. The individual life-paths each contribute to a comprehensive, subjective account of the historical events, the reactions of those recalling them, and the social forces shaping and embracing their personalities. [7]

The database so far has processed the interviews with participants involved in the 1956 revolution. Abstracts of 5-10 pages have been made from the transcripts and these have been incorporated into the database. They are proportionate to the information content in each interview, as summaries in note form. They include the main personal details (place and year of birth, social status of the family, schools, work places etc.) and in greater detail, the historical events in which the interviewee took part, witnessed or heard accounts of, in other words, all the relevant information contained in the interview. After checking the details, the abstracts are incorporated into the finely structured database, where the problem of finding the exact source of information required is reduced to a technical problem with the help of field names. Apart from the technical and administrative data (interviewer, length, accessibility etc.), three main groups of data have been recorded:

1. A description of the interviewee´s life before the revolution (previous life).
2. A description of the interviewee´s life since (after would be a better choice) the revolution (subsequent life).
3. The interviewee´s activity during the revolution (events). [8]

The first two groups of data consist essentially of free text, with a maximum length of 4000 characters, in which it is possible to search for phrases marked. The data was also standardized when the text was introduced into the database. The 1956 events—activities during the revolution—consist of finely structured records (geographical location, exact location, institutions, participants, date, time, side-events etc.) This allow for a very accurate search for memories of specific events. [9]

2.2 Trial records

As research continued, the demand for incorporating other types of document into the database arose. About 13,000 trials took place in the country during the period of reprisals that followed the 1956 Revolution. These concerned more than 25,000 people, including a high proportion of the participants in 'previous' and 'subsequent' political and intellectual life. The trial records provide a great deal of biographical data on those prosecuted (origin, schooling, wealth and so on). Requisite analysis of a well-structured database allows otherwise impossible examinations to be made of participants in the revolution and characteristics of the reprisals. (For instance, what social groups and strata were worst hit? How were foreign observers able to keep track of the sentencing in protracted trials?) The first to be put into the database were the trial records of those who were executed and the data of those who were sentenced to death but the sentence was never carried out. At present, the database contains information on more than two thousand people involved in the `56 trials. [10]

2.3 Chronology

The events are of interest to historians not just in the context of some document, but as distinct chronological events. Therefore, the databases have been compiled from the events as well as the documents. In other words, the events were also incorporated into the database when a period or subject was processed, and links were made, in many cases, to existing records of persons, books, articles, photographs and so on. [11]

2.4 The photographic database

Various audio-visual items (photographs, sound documents and films) appeared by chance with the documents. Apart from their undeniable value as source materials, there is a strong outside demand for these as illustrations for publications of various kinds. Initially, only descriptions of the audio-visual items were incorporated into the database. The systematic processing and digitalization of the material began about three years ago, under various cooperation agreements, along with the establishment of the appropriate links to events and biographies. [12]

3. Establishment of a New Database

The rapid development of the Internet (World Wide Web) was perhaps the main instigating factor when the institute began around 1995, to re-think its procedures for archiving data and developing an adaptable digital archive to meet the changing requirements. [13]

The institute had several databases at that time. In some cases there were structural problems. The records of the same item document found in the different databases were not given in a uniform way. [14]

Some of the tentative requirements for the new system were following:

1. Link all the associated matters found into a single database;
2. store and provide search facilities for long text documents as well;
3. include the audio-visual documents;
4. allow very fine tuning of the accessibility of the database and entitlements to use it;
5. allow some of the data to be public, or even available on the Web;
6. enable statistics to be produced from the database without the user seeing any specific records (to protect personal rights). [15]

With these and many other similar requirements in mind, the institute began in 1996 to develop a databank based on the Oracle software and transferred the data to it. It was found during these operations that there was no established practice for describing, examining or storing the various kinds of document. The

basic concept adopted was to choose the structure that allowed the broadest description to be made, in which some of the data need not be given. [16]

The data did not simply reach the database by being entered within the institute. With the bibliographical data on books and articles, existing descriptions were imported using the standard Marc format for bibliographic exchange before the institute staff subjected them to further processing (linking them with events and persons, for instance). In the absence of a standard, the same procedure was adopted with some descriptions of photographs as well. When cooperating with other institutes and archives, it meant that the data could be entered once and jointly. The sources for compiling them could be provided jointly, rather than the institute having to buy the data. For instance, some of the photo documentation in the database is being developed further in conjunction with the Budapest Archives, in other words, part of it is shared. [17]

The database operated and compiled in this way assists in the work of the "1956 Institute" in two main ways: *by supporting research* and *by assisting publication activities*. [18]

3.1 Database support for research

Researchers can have access to data entered by the institute staff. The database can be used in the same way by everyone, using special recording procedures. Unfortunately, there is a long way to go before the database entirely replaces the card indexes. Not every researcher possesses a computer, and some of the data are still jealously guarded by researchers. However, I think this will change. The stocks of libraries and archives are being digitalized at an increasing rate, so that even if the whole is not available, some description of each item can be found on computer, and often in a restricted form on the Web as well. So the decisive factor in finding the right document increasingly becomes the ability to compose good search questions. Furthermore, the "1956 Institute" is one of several organizations developing intelligent software that is able to search for appropriate information on local databases and on the Internet. Then it can deliver this information to potential users, based on user requirements and the questions that have been formulated. This forms part of the knowledge-processing procedures for the social sciences (which will not spread so fast, of course, as they have in the commercial sphere). This is the direction created by the demand of researchers that the data (from which certain conclusions can be drawn) should be fully accessible, so that statements based on them can be verified or altered as subsequent information emerges. However, this conflicts with the requirement that research findings and data sources should be kept secret. [19]

The Internet allows researchers to use more than one database. Cooperating institutions can compile databases jointly or open their databases to each other´s researchers. These days particularly, the discovery of successive new potential sources of data must be expected. I am thinking here, for instance, of the potential role of corporate archives or of the opening of party and state-security archives in the former socialist countries. The archives of the Gauck Office in

Germany and the Bureau of History in Hungary, for example, are being rapidly explored. [20]

3.2 Database support for publication activities

The second way the new database helps researchers is by giving support for publication activities. The findings of the research done at the "1956 Institute" appear in a variety of publications. Even with a printed publication, it is a great help in preparing the chronology or bibliography, for instance, if the draft text can be compiled from searches in a database that always represents the most up-to-date situation. This applies even more if each item has to be arranged from several points of view. [21]

Meanwhile, several new media have appeared in the information society. These, in my view, do not replace the old media (books and films), but for certain purposes, a CD-ROM or a Web site capable of presenting audio-visual information may be more appropriate than printed publications. This is the case, for instance, with encyclopedic publications containing large volumes of textual data, database handlers are the only means of making a rapid, detailed search. It also applies to works that set out to present a period of history in the most comprehensive possible way (including the arts, way of life, historical characters and so on). [22]

It is important to prepare publications of this kind. Computers and Web usage are part of everyday life for the generation growing up today. We cannot forego the opportunity to convey cultural values and scientific findings through the media they understand the best. [23]

It is also imperative for scientific findings to reach the various levels in the education system. The students need textbooks and teachers need teaching aids. The research institutions also have a responsibility to ensure that information can be transferred rapidly. It is important to have rapid access to the source of authentic, up-to-date information, or to the source with the broadest knowledge on how to obtain it. [24]

The database at the "1956 Institute" serves as the basis for all its publication methods and objectives. It supports various publications, including the Internet series on contemporary Hungarian history since 1945, aimed especially at secondary-school students, and the associated, encyclopedic CD-ROM series for researchers. (The second disc is to appear in 2000 and covers the 1945-56 period.) [25]

The part of the database containing the chronology and the photographic and textual documents will be made accessible to a limited extent. Rather than remaining a closed database, it will gradually develop and alter. The photographic database includes TIFF format photo files for press use, which are not freely accessible on the Internet. The aim in the longer term is to cover some of the mounting expense of maintaining and developing the system by charging fees for

press use of the photo files. This will call for developments in electronic trading that the institute eagerly awaits. [26]

4. The Next Stage

The costs of operating databases and developing them technically and in their content, are appreciable. It is also an expensive undertaking to initiate a new service or prepare a new digital publication, especially if it contains copyright sound or film documents, and the rate of return is slow. I think the demand for such publications will induce enterprising institutions to explore the possibilities of making money using their knowledge or the materials they own. As e-commerce spreads (although many legal details of it are still unclear, especially inter-nationally), it will become possible to make large numbers of comparatively small sales relatively simply and cheaply. This will be especially important for making the right information available at the right time, universally and relatively cheaply. I think many institutions will seize this opportunity. [27]

The criteria of economic efficiency and of substance alike call for links to be made among databases with similar subjects. The users prefer to go to one web site where the answers to a great number of questions and problems can be found, rather than to several sites that answer only some of the questions. Therefore, they are more likely to visit a site if they think that a full answer can be obtained there. In the case of contemporary European history, the histories of nations are linked together in many respects. This applies to the participants in historical events and the events and processes themselves. The research regarding these subjects in the social-science institutes of different countries likewise link together in several ways. From the viewpoint of informatics, much remains to be done here in actually linking up and combining data related in content and concentrating on the many tasks that still need to be done. This provides even more opportunities for scholars analyzing social processes and seeking access to the requisite sources. [28]

This demand for ever more comprehensive links between databases will create many tasks to be solved. One of the question is of standardizing the process of digital archiving of data, which is still at a rudimentary stage. It is high time that standards were set for describing and archiving the various types of digital document. [29]

Smaller nations like Hungary, whose languages are not widely understood, are particularly keen to see databases become multilingual and translatable. Interestingly, the same demand is arising for more widely spoken languages as well, such as German, French and Spanish (English is a special case in this respect), as the number of web users is increasing faster than any language can become a "world language". Online translation of databases (digital textual information) is the only way to ensure international utilization and consistency of content, and the call for it is already increasing rapidly as linguistic techniques develop and aids such as mechanical translation programs and speech-recognition programs become available. The task of introducing these aids into

the social sciences has to be tackled. Unfortunately, it seems as if it will be more difficult to arrive at requisite means in this field, because historical texts are much less specific. The development of them will be more costly, because the linguistic structures involved are less standard, for instance, than they are in legal texts. Another relevant factor, at least for the time being, is that the solvent demand for translations of this kind will remain smaller than for legal texts. [30]

There remains plenty to do in linking databases and making it possible to link them, and in finding the funds to build them up, for I have not mentioned the marketing aspects at all. There is a need for real cooperation while each institution continues to develop its own system according to the requirements and opportunities. [31]

Author

Zoltán LUX

Contact:

Zoltán Lux

Institute for History of the 1956 Hungarian Revolution
H-1074 Budapest, Dohány u. 74.

Phone: 36-1-322-5228
Fax: 36-1-322-3084

E-mail: luxz@helka.iif.hu
URL: http://www.rev.hu

Citation